

# いつ誰が何を話したか？

～複数人会話シーン分析技術の進展と利用イメージ～

## どんな研究？

人間同士の自然なコミュニケーションを認識し、状況を理解することでコミュニケーションを分析・支援する技術が求められています。私たちは、実際の複数人の会話から「いつ、誰が、誰に、何を話したか」を高い精度で認識する技術を研究しています。

## どこが凄い？

実環境下の会話における音声認識は、各話者の声が雑音や他の人の声などの影響を受け、とても困難です。これに対し、私たちが開発した雑音除去法や音声/言語モデル学習法などを用いることで、卓上マイク利用の場合にもヘッドセットマイク利用の場合に匹敵する高い音声認識精度を実現しました。

## どんな風に役立つ？

会話音声を高精度で認識できれば、過去の会議の検索や、議事録の作成が非常に容易になります。さらにリアルタイム版を利用すれば、参加者が知らない用語や聞き逃した内容を会話中に検索したり、遠隔地の傍聴者が高臨場感で会話に参加することも可能となります。

## 会話シーン分析技術の最近の進展についてご紹介

### 複数人の会話シーン

遠隔マイクで収録される自由会話音声  
(残響・背景雑音・同時発話を含む)

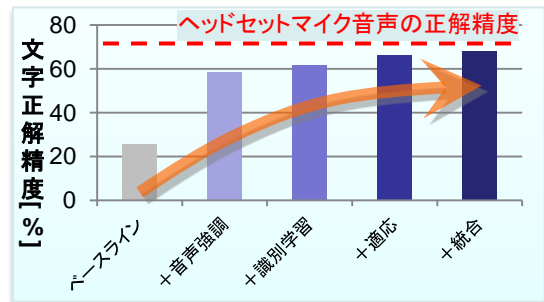
➡ 認識・理解・支援する技術



### ① 複数人による遠隔・自由発話音声の認識精度向上

独自技術:

- ◆ 音声モデルの識別学習・適応技術
- ◆ 各種マルチチャネル音声強調処理の統合により、ヘッドセットマイク音声に匹敵する認識精度を達成



### ② 会話支援タブレットの進化

「いつ誰が何を話したか」を分析し、会話をモニタリング

- ◆ 発言へのメモ付与や用語の検索
- ◆ 過去の会話の検索・再生
- ◆ 日本語・英語に対応

会話の効率化・活性化、過去の会議情報の整理・活用を支援



### 関連文献

- [1] T. Hori, S. Araki, T. Yoshioka, M. Fujimoto, S. Watanabe, T. Oba, A. Ogawa, K. Otsuka, D. Mikami, K. Kinoshita, T. Nakatani, A. Nakamura, and J. Yamato, "Low-latency real-time meeting recognition and understanding using distant microphones and omnidirectional camera," IEEE Trans. ASLP, vol. 20, no. 2, pp. 499-513, Feb. 2012.
- [2] 堀, 荒木, 小川, ソウデン, デルクロア, 吉岡, 大庭, 藤本, 木下, 久保, ハム, 渡部, 中谷, 中村, "会話分析タスクにおける複数人自由会話の遠隔発話音声認識の評価", 日本音響学会講演論文集, 3-P-5, March 2012.

### 連絡先

堀貴明 (Takaaki Hori) メディア情報研究部 信号処理研究グループ  
E-mail : hori.t{at}lab.ntt.co.jp ( {at} の部分を @ に置き換えてください)